不同地理种群二化螟 Ty3/*gypsy* 反转座子天冬氨酰蛋白酶(AP)基因序列的克隆与分析

李晓欢[1,2,#], 罗光华[1,#], 韩召军[2], 方继朝[1,2,*]

(1. 江苏省农业科学院植物保护研究所,南京 210014; 2. 南京农业大学植物保护学院,南京 210095)

摘要:【目的】Ty3/*gypsy* 反转座子是广泛存在于生物体内的一类反转座子。反转座子上的天冬氨酰蛋白酶(aspartic protease, AP)基因是反转座子发生转座所需的一个重要基因。但由于该基因家族成员间变异较大,较难利用简并引物克隆得到该基因,所以对该基因家族成员的研究很少。【方法】本研究采用 PCR 方法克隆了二化螟 Ty3/*gypsy* 反转座子的 AP 基因序列,并对其序列特征和地理种群变异进行了分析。【结果】克隆获得的二化螟 *Chilo suppressalis* (Walker) Ty3/*gypsy* 反转座子中的 AP 基因具有独立的开放阅读框(open reading frame, ORF),长 528 bp,编码的蛋白含 175 个氨基酸残基(GenBank 登录号:KF886014)。Conserved Domain Search 在线工具分析显示,该蛋白中含有一个特异的 Asp_protease_2 保守功能域。从 7 个二化螟不同地理种群中共克隆获得 70 份 AP 基因拷贝。对同一基因座位上的 AP 序列多重比对分析,发现共存在 46 处碱基替换,其中碱基转换(transition)有 31 处,碱基颠换(transversion)有 15 处,70 份拷贝中有 69 份拷贝是完整的 ORF,能编码完整的蛋白。从碱基替换形式看,A→G 的变异形式出现最多,有 15 处;其次是 T→C 的变异形式,有 11 处;其余的变异形式都很少。对比这 7 个不同地理种群,没有发现碱基的替换存在明显的地理区划差异。【结论】碱基的替换形式与二化螟所处的地理区域无明显相关性。本研究对于认识反转座子序列的变异特点有所帮助。

关键词: 二化螟; 反转座子; Ty3/*gypsy*; 天冬氨酰蛋白酶; 基因克隆; 序列分析

中图分类号: Q966 **文献标识码:** A **文章编号:** 0454-6296(2014)05-0530-08

Molecular cloning and analysis of aspartic protease (AP) gene in Ty3/*gypsy* retrotransposon in different geographical populations of *Chilo suppressalis* (Lepidoptera: Pyralidae) in China

LI Xiao-Huan[1,2,#], LUO Guang-Hua[1,#], HAN Zhao-Jun[2], FANG Ji-Chao[1,2,*] (1. Institute of Plant Protection, Jiangsu Academy of Agricultural Sciences, Nanjing 210014, China; 2. College of Plant Protection, Nanjing Agricultural University, Nanjing 210095, China)

Abstract:【Aim】Ty3/*gypsy* is a retrotransposon that is widespread in many organisms. Aspartic protease (AP) gene is important for the retrotransposon transposition. Since the AP gene is difficult to be cloned based on degenerate primers because the members of AP family have low similarities, there are few researches about AP genes.【Methods】The AP gene in Ty3/*gypsy* retrotransposon was cloned from *Chilo suppressalis* (Walker) with PCR method, and its sequence characteristics and variations in the Chinese populations were analyzed.【Results】The AP gene has a single open reading frame (ORF) of 528 bp encoding a protease of 175 amino acids (GenBank accession no.: KF886014). The Conserved Domain Search analysis showed that there is a conserved Asp_protease_2 domain in this AP protein. Seventy copies of AP gene were cloned from 7 field populations. Multiple sequences alignment revealed that there are 46 single-base substitutions including 31 transitions and 15 transversions. Sixty-nine of 70 copies have intact ORF. The base changes of A→G (15) and T→C (11) occur more frequently than the others.【Conclusion】There is no significant relationship between the patterns of base substitutions and the distribution area of *C. suppressalis*. This study will be helpful to investigate the variation features of retrotransposon.

基金项目: 国家水稻产业技术体系项目(Cars-001-25); 国家自然科学基金项目(31201505); 江苏省农业科技自主创新资金项目 CX(13)5025

作者简介: 李晓欢, 女, 1989 年生, 辽宁辽阳人, 硕士研究生, 研究方向为昆虫分子生物学, E-mail: xhli2012@163.com; 罗光华, 男, 1982 年生, 江苏溧阳人, 博士, 副研究员, 研究方向为粮食作物虫害, E-mail: luogh_cn@163.com

[#]共同第一作者 Authors with equal contribution

[*]通讯作者 Corresponding author, E-mail: fangjc@jaas.ac.cn

收稿日期 Received: 2013-11-28; 接受日期 Accepted: 2014-04-11

Key words: *Chilo suppressalis*; retrotransposon; Ty3/*gypsy*; aspartic protease (AP); gene cloning; sequence analysis

转座子是一段 DNA 序列,能够不依靠同源重组而在宿主基因组内移动。起初转座子被认为是垃圾 DNA 序列,在宿主体内无生物学功能(Doolittle and Sapienza, 1980),但现在已经明确,转座子是宿主基因组的重要组成部分,是宿主进化的重要内在动力之一(Kidwell and Lisch, 2000; Kidwell, 2002; Kazazian, 2004)。

按照转座机制和序列特征可将转座子分为两大类,第一大类是 RNA 型转座子,也称为反转座子,通过"复制-粘贴"的方式进行转座,该类转座子每发生一次转座则在宿主基因组中多一份自身的拷贝,因此转座活性越强,其在宿主染色体内的拷贝数就越多,由此增加了宿主基因组的容量(Finnegan, 1997)。第二类是 DNA 型转座子,通过"剪切-粘贴"的方式进行转座,在转座酶的作用下,转座子自身直接从宿主基因组的一个位点转移到另一个位点,不会对宿主基因组的大小产生影响,但能导致宿主核酸序列的转移和序列位置的变化(Feschotte and Pritham, 2007)。

反转座子是种类最多、分布最广的转座子类型。反转座子根据 5′和 3′端末端重复序列的有无,分为长末端重复(long terminal repeat, LTR)反转座子和非长末端重复(non-long terminal repeat, non-LTR)反转座子(Bennetzen, 1996)。LTR 反转座子又可以分为 5 个超家族(superfamily): Ty1/*Copia* (Pseudoviridae), Ty3/*Gypsy* (Metaviridae), Bel/Pao, retroviruses 以及 ERV (endogenous retroviruses)。直到现在,后 3 个超家族成员只在后生动物体内发现,而 Ty1/*Copia* 和 Ty3/*Gypsy* 两个超家族成员在所有真核生物体内都有发现(Muszewska *et al.*, 2011)。Ty1/*Copia* 和 Ty3/*Gypsy* 的区别就在于所编码的多聚蛋白中整合酶(integrase, INT)的位置差异(Kumar and Bennetzen, 1999; Muszewska *et al.*, 2011)。

LTR 反转座子中都含有一段编码天冬氨酰蛋白酶(aspartic protease, AP)的序列。该蛋白酶属于逆转录病毒蛋白酶类似天冬酰胺蛋白酶家族(retroviral pepsin-like aspartate protease family),该家族一般含有 180～230 个氨基酸残基,其 C-末端相对较保守,N-末端变化较大。天冬酰胺蛋白酶家族成员之间相似度较低,所以较难通过同源克隆获得该蛋白酶家族成员。现在已报道的该蛋白酶家族成员大多数是借助于生物信息学工具从各种类型的数据中发掘得到的(Llorens *et al.*, 2009)。LTR 反转座子所编码的天冬氨酰蛋白酶(AP),其在反转座子内部可以是一个独立的 ORF,也可以与反转录酶(reverse transcriptase, RT)、RNA 酶 H(RNase H, RH)以及整合酶(INT)共存在一个 ORF 中(Flavell *et al.*, 1997; Havecker *et al.*, 2004; Muszewska *et al.*, 2011)。在 LTR 反转座子中,天冬氨酰蛋白酶(AP)的作用是在反转座子转座的过程中负责将反转座子表达的多聚体蛋白切割成能行使功能的成熟蛋白产物(Muszewska *et al.*, 2011)。因此,对 LTR 反转座子 AP 的研究可以帮助了解 LTR 反转座子的变异规律,进而为深入解析反转座子的增殖扩散规律及沉默失活机制提供理论依据。

本研究从水稻的重要害虫二化螟 *Chilo suppressalis* (Walker)体内克隆 Ty3/*gypsy* 反转座子的 AP 序列,利用相关生物信息学工具对该 AP 序列进行分析,从而认识其重要的分子生物学结构与特征。此外,比较分析了不同二化螟地理种群中该 AP 序列的变异,以此探讨宿主所处的不同地理区域与内源性转座子序列的变异的关系,以期为揭示二化螟的 Ty3/*gypsy* 反转座子遗传进化规律提供有用的信息。

1　材料与方法

1.1　不同地区二化螟种群的采集

于 2010 年 12 月至 2011 年 2 月沿长江流域,分别从四川德阳(104°20′36″E, 31°7′56″N)、重庆江津(106°15′54″E,29°8′28″N)、湖北荆州(112°20′29″E, 30°16′48″N)、湖南湘阴(112°54′28″E,28°38′45″N)、江西瑞昌(115°11′22″E,28°43′38″N)、安徽和县(118°21′25″E,31°43′3″N)和浙江鄞县(121°31′43″E,29°48′26″N)的水稻田内采集二化螟幼虫,采集的幼虫经饥饿处理后浸泡于无水乙醇中,置于 －80℃ 冰箱冻存备用。

1.2　基因组 DNA 的提取

利用 Axygen 公司的基因组提取试剂盒(Axygen 生物科技有限公司,杭州)提取二化螟的基因组 DNA(gDNA)。提取 gDNA 时取单头二化螟,用吸水纸将虫体表面的乙醇吸干后用液氮速冻虫体,再用

研磨棒研磨至粉末状,之后按照基因组提取试剂盒的使用说明书进行操作。提取的 gDNA 作为后续 PCR 的模板,于 -20℃ 保存。

1.3 总 RNA 提取及第一链 cDNA 模板的制备

每个地理种群随机挑选二化螟 3 龄幼虫 2 头,提取总 RNA。总 RNA 的提取使用 SV Total RNA Isolation Kit 试剂盒(Promega 公司,美国),具体操作步骤依照说明书。提取的总 RNA 琼脂糖凝胶电泳和紫外分光光度计检测质量后,用 PrimeScript II 1st Strand cDNA Synthesis Kit 反转录试剂盒(大连宝生物公司,大连)反转录合成第一链 cDNA。反转录产物于 -20℃保存备用

1.4 二化螟 Ty3/*gypsy* 反转座子 AP 序列的克隆

根据我们早先克隆的二化螟 Ty3/*gypsy* 反转座子序列 Csu-Ty3(GenBank 登录号:KJ191261),设计特异引物克隆该 Ty3/*gypsy* 反转座子的 AP 序列。所设计的引物为:Ty3-APF1:5′-GAGGAAGCGATAAGGATCA-3′,Ty3-APR1: 5′-TGAAGCTTGCCTACTTAACAGAG-3′。引物位于 Ty3/*gypsy* 反转座子序列中 AP 结构域两端旁侧。以二化螟 gDNA 为模板,使用该对引物克隆 Ty3/*gypsy* 反转座子的 AP 序列。利用 TaKaRa 公司(大连宝生物公司)的 *LA Taq* 酶进行扩增,共 25 μL 反应体系:含 2.5 μL 的 10 × PCR Buffer, 2.5 μL 的浓度为 25 mmol/L 的 $MgCl_2$、2 μL 的浓度均为 2.5 mmol/L 的 dNTPs、0.25 μL 的浓度为 5 U/μL 的 *LA Taq* 酶、浓度均为 10 μmol/L 的上下游引物各 1 μL。PCR 扩增程序为:95℃预变性 3 min;94℃变性 30 s,56℃退火 30 s,72℃延伸 45 s,进行 32 个循环;最后 72℃延伸 10 min。同时,以二化螟 cDNA 为模板,引物、反应体系及扩增程序均同上。扩增 Ty3/*gypsy* 反转座子 AP 结构域的 cDNA 序列。

PCR 产物用 1.5% 的琼脂糖凝胶进行电泳检测,溴化乙锭(EB)染色后将潜在的目标条带切割下来,利用 DNA 凝胶回收试剂盒(天根生物技术有限公司,北京)回收 PCR 产物。回收的 DNA 产物连接至 pEASY-T3 克隆载体(北京全式金生物技术有限公司,北京),转化后挑取单克隆摇菌培养并测序。

1.5 二化螟 Ty3/*gypsy* 反转座子 AP 的分析

将该 AP 的核酸序列在 NCBI 的 Nucleotide Collection (nr/nt)、Whole-Genome Shotgun Contigs (wgs) 数据库(http://www.ncbi.nlm.nih.gov/blast/)中利用 BLASTN 工具进行同源搜索;利用 NCBI 的 Conserved Domain Search 在线分析工具(http://www.ncbi.nlm.nih.gov/Structure/cdd/wrpsb.cgi)对该蛋白进行保守结构域预测分析;使用在线分析工具 SignalP 4.1 Server(http://www.cbs.dtu.dk/services/SignalP/)预测该 AP 蛋白的信号肽序列;使用 TMpred 跨膜蛋白预测在线分析工具(http://www.ch.embnet.org/software/TMPRED_form.html)对该 AP 蛋白进行跨膜区分析;利用 PredictProtein 在线工具(https://www.predictprotein.org/)对 AP 蛋白的二级结构等进行预测分析。

1.6 二化螟不同地理种群的 Ty3/*gypsy* 反转座子 AP 基因的变异分析

为比较二化螟 Ty3/*gypsy* 反转座子同一拷贝中 AP 基因的变异,首先基于我们前期克隆获得的该 Ty3/*gypsy* 反转座子的旁侧序列设计特异引物(Ty3-flkF:5′-ATCCATTGGCGTGCTGTTAT-3′和 Ty3-flkR:5′-TACACAATAGAGCGTTCTGCTAC-3′), 进行 Flank-PCR,扩增二化螟不同地理种群中同一基因座位上 Ty3/*gypsy* 反转座子拷贝。利用 TaKaRa 公司(大连宝生物公司)的 *LA Taq* 酶进行扩增。Flank-PCR 的扩增程序为:95℃预变性 3 min;94℃变性 30 s,58℃退火 30 s,72℃延伸 5 min 20 s,进行 30 个循环;最后 72℃延伸 10 min。扩增产物回收后作为后续 PCR 的模板。再利用前述的特异引物(Ty3-APF1:5′-GAGGAAGCGATAAGGATCA-3′和 Ty3-APR1:5′-TGAAGCTTGCCTACTTAACAGAG-3′),以 Flank-PCR 的回收产物为模板,克隆 Ty3/*gypsy* 反转座子同一拷贝上的 AP 序列。PCR 扩增程序为:95℃预变性 3 min;94℃变性 30 s,56℃退火 30 s,72℃延伸 45 s,进行 32 个循环;最后 72℃延伸 10 min。将目标条带切割下来后回收、连接、转化、挑取单克隆、摇菌培养并测序,使用的试剂与操作方法同上。每个地理种群中,随机挑选 10 头个体,单头提取基因组 DNA,作为 Flank-PCR 的模板。

利用 MEGA 4 软件对从不同地理种群中克隆得到的 AP 基因核酸序列进行分子系统进化分析,采用最大简约法(maximum parsimony, MP)1 000 次重复构建分子系统进化树。

2 结果与分析

2.1 二化螟 Ty3/*gypsy* 反转座子中 AP 基因的序列与结构预测

二化螟 Ty3/*gypsy* 反转座子的 AP 基因核苷酸序列长 528 bp,在二化螟 Ty3/*gypsy* 反转座子中是

一个独立的 ORF,拥有自身的起始密码子和终止密码子(研究显示,在有的反转座子中,AP 基因序列结构会被包含在整个 *pol* 多聚体基因中,不具有自身的起始密码子和终止密码子),其编码含 175 个氨基酸残基的蛋白(图 1)。以 cDNA 为模板进行的 PCR 扩增,同样能获得相同的序列。



图 1　二化螟 Ty3/*gypsy* 反转座子上 AP 基因的核酸序列以及推导的蛋白序列

Fig. 1　Nucleotide and putative amino acid sequences of AP gene in retrotransposon Ty3/*gypsy* from *Chilo suppressalis*

星号表示终止密码子。The asterisk indicates the stop codon.

将该 AP 核酸序列在 NCBI 的 Nucleotide collection (nr/nt)数据库中进行 BLSATN 分析,当选择 Highly similar sequences (megablast)程序时搜索不到任何同源序列;当选择 More dissimilar sequences (discontiguous megablast)程序时仅搜索到一条序列,所显示的 E 值为 3.3,而且两者之间也只有 32 bp 是匹配的;当选择 Somewhat similar sequences(blastn)程序时能搜索出很多序列,但这些序列都是一些物种的基因组或其他组学的序列,尚未被进一步的研究确认,此外所有这些搜寻结果中 E 值最低为 0.002。将该 AP 核酸序列在 NCBI 的 Whole-Genome Shotgun contigs (wgs)数据库中进行 BLSATN 分析,获得的结果与上述结果相似。将该 AP 蛋白序列进行 BLASTP 分析,选择的搜索程序为 BLASTP(protein-protein BLAST),搜索出了 102 个结果(blast hits)。这些 hits 中 E 值最小的为 7e-18,最大的为 0.14。其中绝大部分序列都被列为未知功能蛋白,少数序列归属于反转座子的序列。

利用 NCBI 的 Conserved Domain Search 在线分析工具进行分析,结果显示该蛋白中含有一个特异的 Asp_protease_2 保守功能域,在该功能域中有一个 catalytic motif(图 2)。利用在线分析工具 SignalP 4.1 Server 进行信号肽预测,没有发现该蛋白中存在信号肽结构。利用 TMpred 在线分析工具分析跨膜区,没有发现跨膜区结构。

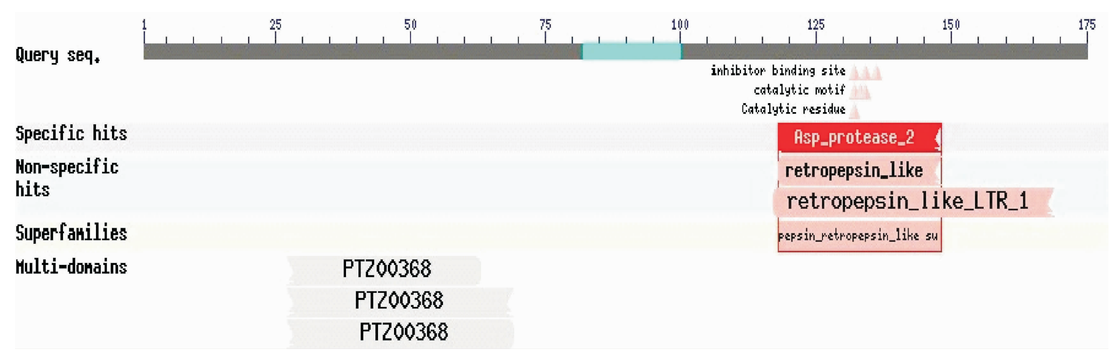


图 2　二化螟反转座子 Ty3/*gypsy* 上 AP 基因蛋白序列保守结构域预测

Fig. 2　Prediction of conserved domains of coding products of AP gene in retrotransposon Ty3/*gypsy* from *Chilo suppressalis*

Inhibitor binding sites：抑制结合位点；Catalytic motif：催化功能域基序；Catalytic residue：催化残基；Specific hits：特异锚定匹配；Non-specific hits：非特异锚定匹配；Superfamilies：超家族；Multi-domains：多重结构域.

对 AP 蛋白利用 PredictProtein 在线工具分析显示,在 AP 蛋白二级结构中 5.7% 的氨基酸残基参与形成螺旋(helix),12.6% 的氨基酸残基参与形成片层(strand),其余的氨基酸残基则参与形成环(loop),其中螺旋所处的位置约在第 145 位氨基酸附近,片层所处位置约在第 128 位氨基酸附近(图 3)。Gene Ontology(GO)分析中可信度最高的(38%)是将该 AP 蛋白的分子生物学功能归属于 Protein Binding 这一类中,蛋白结合位点(Protein-Protein Binding)分析显示该 AP 蛋白中散布着很多个潜在的蛋白结合位点。

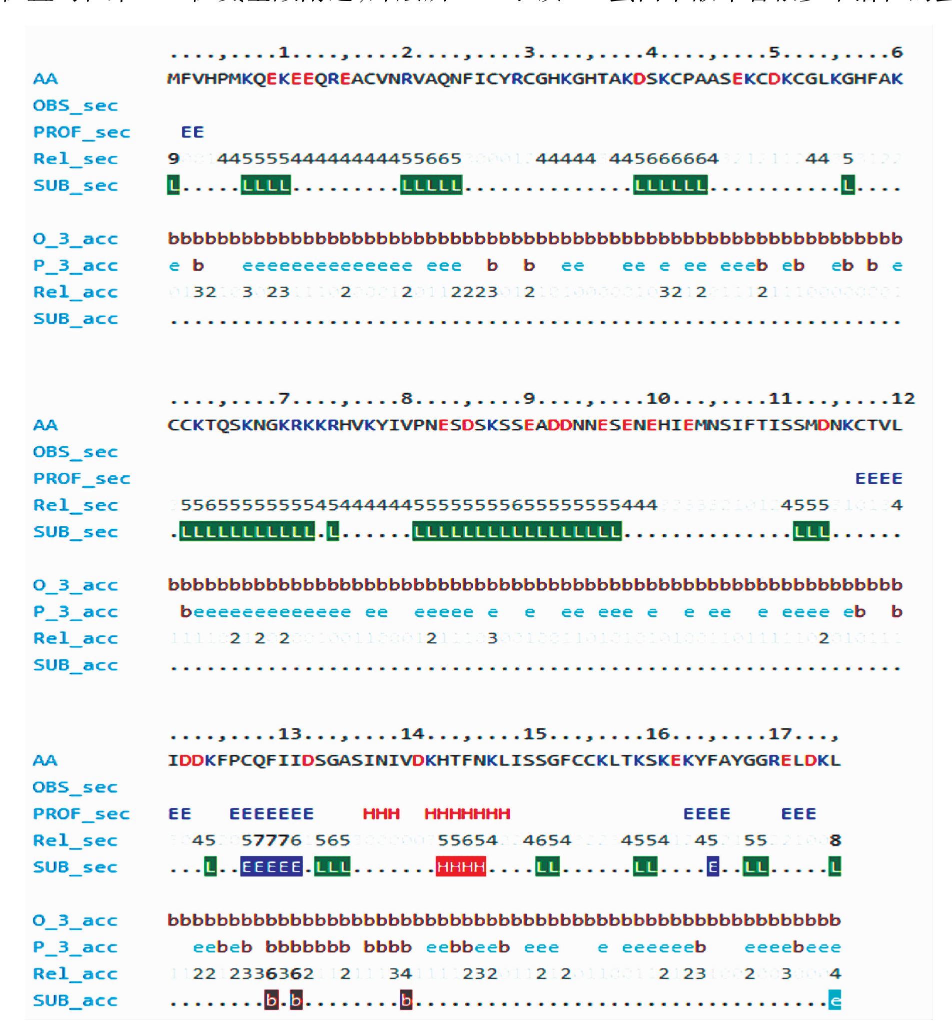


图 3 二化螟反转座子 Ty3/*gypsy* 上 AP 基因编码蛋白的二级结构预测

Fig. 3 Prediction of the secondary structure of the coded protein of AP gene in retrotransposon Ty3/*gypsy* from *Chilo suppressalis*

E: 片层 Strand; H: 螺旋 Helix; L: 环 Loop. SUB: 在置信度范围内校正后的值 Correction value in confidence intervals. 预测结果以 SUB 值为准。Predicted results are based on the SUB values.

2.2 不同二化螟地理种群中同 AP 的变异分析

在 7 个不同地理种群的二化螟体内均克隆到了 Ty3/*gypsy* 的 AP 序列,共获得 70 份拷贝。在这 70 份拷贝中,只有 1 份拷贝(江津种群中的第 3 号样)存在一段连续的 297 bp 的缺失,其余拷贝都没有碱基的缺失,也没有碱基的插入。此外,所有这些没有碱基插入和缺失的拷贝都能够完全翻译通,即不存在碱基的无义突变(nonsense mutation)。这些拷贝相互之间保持着极高的相似度(除了江津种群中的第 3 号拷贝,因其有较大片段的缺失),核酸水平的相似度≥99%,蛋白水平的相似度≥98%。这表明,在 7 个跨度较大的地理区域之间,作为非远距离迁飞的二化螟种群,它们同一基因座位上的内源性 Ty3 反转座子序列没有出现显著的差异。基于 AP 基因核酸序列构建的分子系统进化树(未列出),结果显示不同地理种群中的 AP 基因拷贝没有明显的聚类,也表明该基因序列在不同地理种群中变异较小。

序列多重比对显示,共存在 46 处碱基替换,其中碱基转换(transition)有 31 处;碱基颠换(transversion)有 15 处(表 1)。对于碱基的变异,从三联体密码子的位置上来分析,密码子第 3 位出现变异的次数最多,有 20 处;第 1 位和第 2 位出现变

异的次数相同,均为 13 处。碱基变异导致的氨基酸同义突变(synonymous mutation)有 18 处,且均出现在密码子的第 3 位;碱基变异导致的氨基酸错义突变(missense mutation)有 28 处,有 13 处均出现在密码子第 1 位和第 2 位,还有 2 处出现在密码子第 3 位(表 1)。

表 1　二化螟反转座子 Ty3/*gypsy* 上 AP 基因序列的变异情况

Table 1　Substitution mutations of AP genes in retrotransposon Ty3/*gypsy* from *Chilo suppressalis*

	碱基转换 Transition			碱基颠换 Transversion		
	密码子中位置 Position in codon			密码子中位置 Position in codon		
	第 1 位 Position 1	第 2 位 Position 2	第 3 位 Position 3	第 1 位 Position 1	第 2 位 Position 2	第 3 位 Position 3
不同位置上的变异个数 Number of substitutions	11	6	14	2	7	6
同义突变数 Number of synonymous mutations	0	0	14	0	0	4
错义突变数 Number of missense mutations	11	6	0	2	7	2
无义突变数 Number of nonsense mutations	0	0	0	0	0	0
碱基变异总数 Total number of base substitutions		31			15	

从变异形式上分析碱基的变异,A→G 的变异形式出现最多,有 15 处;其次是 T→C 的变异形式,有 11 处。其余的变异形式都很少,其中未发现 C→T 和 C→A 的变异形式(表 2)。对比这 7 个不同地理种群,没有发现碱基的替换存在明显的地理区划差异,这显示该转座子的序列变异与宿主二化螟所处的地理区域无明显相关性。

表 2　二化螟反转座子 Ty3/*gypsy* 上 AP 基因序列间单碱基替换形式

Table 2　Single-base substitution patterns of AP genes in retrotransposon Ty3/*gypsy* from *Chilo suppressalis*

变异类型 Substitution type	形式 Forms	变异个数 Number of substitutions	密码子中位置 Position in codon		
			第 1 位 Position 1	第 2 位 Position 2	第 3 位 Position 3
碱基转换 Transition	A→G	15	7	2	6
	G→A	5	2	0	3
	T→C	11	2	4	5
	C→T	0	0	0	0
碱基颠换 Transversion	A→T	3	0	1	2
	T→A	3	0	0	3
	A→C	1	0	1	0
	C→A	0	0	0	0
	G→T	2	0	2	0
	T→G	1	0	0	1
	G→C	2	2	0	0
	C→G	3	0	3	0

3　讨论

本研究首次克隆获得二化螟 Ty3/*gypsy* 反转座子的天冬氨酰蛋白酶(AP)基因序列。该 AP 基因的 BLASTN/P 分析均未找到相似(同源)序列,或只有个别 E 值较高而且尚未研究确证的序列片段,这表明目前对于反转座 AP 基因序列的研究较少。相关报道认为天冬酰胺蛋白酶家族成员之间相似度较低(Llorens *et al*., 2009; Muszewska *et al*., 2011),这可能是我们进行 BLAST 分析找不到相似(同源)序列的原因。由此,意味着对 AP 序列的同源克隆较困难,这或许也是目前 AP 序列研究较少的原因。鉴于此,本研究对我们更系统地认识反转座子的 AP 基因序列及其结构特点具有重要的参考价值。

虽然从 BLAST 分析结果中来看,并不能很明确所克隆的序列是 AP 基因,但 Conserved Domain Search 分析和 GO 分析所提供的信息使我们相信所克隆的序列是 AP 基因。AP 基因编码蛋白的功能主要是切割多聚蛋白体,也即蛋白与蛋白间的相互作用,因而其必定含有蛋白结合位点,对序列的生物信息学分析也证实了这一点。此外该序列是从 Ty3/*gypsy* 反转座子中克隆获得,并且其在 Ty3/*gypsy* 反转座子中的位置符合已报道的 AP 基因的特征,结合上述这些信息,我们确认所克隆的序列就是 AP 基因。进一步对该 AP 基因的蛋白二级结构分析,蛋白序列中没有信号肽结构,也不存在跨膜区,表明 AP 蛋白是在胞内行使功能的。

在二化螟不同地理种群中,克隆得到 70 份 AP 序列拷贝。多重序列比对结果显示碱基转换发生的次数明显多于碱基颠换发生的次数,相同的现象在

人的癌基因变异中也存在(Prathima, 2012)。从碱基分子的空间构像上分析,嘌呤碱基均有两个环,嘧啶碱基都只有一个环,个体大小相似的碱基之间发生替换肯定比一大一小之间发生替换容易的多,所以碱基转换发生的几率会大于碱基颠换发生的几率。由于密码子的兼并性,所以密码子第 3 位发生的变异的几率必定大于其他两个位置,我们的结果以及前人研究的结果都证实了这一点(Prathima, 2012)。对于碱基替换的形式,在癌细胞基因的碱基替换中 C→T 和 G→A 是出现最多的两类,被认为是与碱基的甲基化有关(Bird, 1980; Arnheim and Calabrese, 2009; Prathima, 2012)。而我们的结果则显示在反转座子的 AP 序列中 A→G 和 T→C 是出现最多的两类碱基替换形式。在癌基因和转座子基因中为何存在这种差异,需要更深入的研究。这是否是转座子序列变异的普遍规律,也值得我们进一步研究。有研究表明,内源性转座子的序列甲基化是转座子保持失活状态的一个重要原因(Miura *et al.*, 2001; Weil and Martienssen, 2008)。我们的分析结果显示该反转座子序列倾向于向"G/C"的变异,这可能会增加潜在的甲基化位点,从而使转座子更容易受到宿主的甲基化调控。对此,我们正在更细致地研究。

转座子是宿主进化的重要内在动力之一,转座子能够通过改变宿主的某些重要功能基因或宿主遗传多样性进而提高宿主的进化适应能力(Beguiristain *et al.*, 2001; Bohne *et al.*, 2008; Oliver and Greene, 2009)。在某些特殊情况下,宿主能够解除对其体内转座子转座的调控,从而提高宿主自身的环境适应能力,因此宿主体内保留具有潜在活性的转座子对宿主具有重要的生物学意义(Beguiristain *et al.*, 2001; Dooner and Weil, 2007; Cam *et al.*, 2008)。但在大部分情况下,转座子的转座会影响宿主的某些正常功能基因或者导致宿主染色体的异常重组等,因此转座子的转座受到宿主的严格调控(Beguiristain *et al.*, 2001; Dooner and Weil, 2007; Bohne *et al.*, 2008)。由此可以看出,转座子与宿主存在于某个复杂的系统中,这一定程度上是因为转座子与宿主基因组中其他基因之间有着复杂的相互关系。从宏观上看,如果将整个宿主基因组看成是一个生态系统,其每个组分(如一个基因或一段核酸片段)都会与其他组分相互影响,形成一定的互惠关系,共生关系,也可能是寄生关系(Brookfield, 2005;张博等,2009)。当然,也可能存在竞争关系。

二化螟作为一种重要的农业害虫,虽然无法长距离迁徙,但却广泛分布。我们所选择的 7 个二化螟地理种群,在地理位置上相隔较远,且地理环境与气候条件差异明显,但都属于长江流域范围。Meng 等(2008)基于微卫星和线粒体 DNA 将我国二化螟种群划分成 3 大分支:华中分支(central China, CC)、华北加东北分支(northern plus northeastern China, NN)和西南分支(southwestern China, SW)。对照此,我们所选的德阳、江津位于 SW 分支内;荆州位于 NN 分支内;湘阴、瑞昌、和县和鄞县位于 CC 分支内。研究认为这 3 个分支之间基因交流较少,而我们从这 7 个不同地理种群中共克隆得到 70 份 AP 序列拷贝,有着极高的相似性,而且其中 69 份拷贝均具有完整的 ORF。这表明,尽管宿主二化螟在不同地理区域中需要适应各自所处的地理环境,但它们体内的反转座子序列(AP 序列)却在长期的进化过程中保持着完整的序列结构。我们推测这其中可能有两个原因:第一,本实验所研究的该 AP 序列及其所在的反转座子与二化螟地理种群的分化或适应进化可能没有相关性。第二,或许是该 AP 序列及其所在的反转座子在与宿主二化螟协同进化的过程中对宿主进化具有重要的生物学意义,从而使得自身能够完整的在宿主体内保存下来。因为在多数情况下,宿主对转座子的调控往往都使转座子丧失完整性而失去转座活性(Wang *et al.*, 2008; Luo *et al.*, 2011)。此前,我们在二化螟体内克隆得到一类 DNA 型转座子(*CsuPLE*)的多份拷贝,并且这些拷贝大多数仍具有完整的序列结构(罗光华等, 2012),也存在着上述相似的情况。综上,我们认为二化螟在长期的进化过程中可能受到其体内多种类型转座子的影响,二化螟体内多种转座子的存在使其拥有更丰富的遗传多样性。我们正在试图寻找到转座子引起的二化螟种群差异或表型变化,为揭示转座子在二化螟种群进化过程中起到的重要作用提供确切的理论证据。

参考文献(References)

Arnheim N, Calabrese P, 2009. Understanding what determines the frequency and pattern of human germline mutation. *Nature Reviews Genetics*, 10(7): 478-488.

Beguiristain T, Grandbastien MA, Puigdomenech P, Casacuberta JM, 2001. Three Tnt1 subfamilies show different stress-associated patterns of expression in tobacco. Consequences for retrotransposon control and evolution in plants. *Plant Physiology*, 127: 212-221.

Bennetzen JL, 1996. The contributions of retroelements to plant genome organization, function and evolution. *Trends in Microbiology*, 4(9): 347-353.

Bird AP, 1980. DNA methylation and the frequency of CpG in animal DNA. *Nucleic Acids Research*, 8: 1499-1504.

Bohne A, Brunet F, Galiana-Arnoux D, Schultheis C, Volff JN, 2008. Transposable elements as drivers of genomic and biological diversity in vertebrates. *Chromosome Research*, 16: 203-215.

Brookfield JFY, 2005. The ecology of the genome-mobile DNA elements and their hosts. *Nature Reviews Genetics*, 6: 128-136.

Cam HP, Noma K, Ebina H, Levin HL, Grewal SI, 2008. Host genome surveillance for retrotransposons by transposon-derived proteins. *Nature*, 451: 431-437.

Doolittle WF, Sapienza C, 1980. Selfish gene, the phenotype paradigm and genome evolution. *Nature*, 284: 601-603.

Dooner HK, Weil CF, 2007. Give-and-take: interactions between DNA transposons and their host plant genomes. *Current Opinion in Genetics & Development*, 17: 486-492.

Feschotte C, Pritham EJ, 2007. DNA transposons and the evolution of eukaryotic genomes. *Annual Review of Genetics*, 41: 331-368.

Finnegan DJ, 1997. Transposable elements: how non-LTR retrotransposons do it. *Current Biology*, 7(4): R245-R248.

Flavell A, Pearce SR, Heslop-Harrison P, Kumar A, 1997. The evolution of Ty1-*copia* group retrotransposons in eukaryote genomes. *Genetica*, 100: 185-195.

Havecker ER, Gao X, Voytas DF, 2004. The diversity of LTR retrotransposons. *Genome Biology*, 5: 225.

Kazazian HHJr, 2004. Mobile elements: drivers of genome evolution. *Science*, 303(5664): 1626-1632.

Kidwell MG, 2002. Transposable elements and the evolution of genome size in eukaryotes. *Genetica*, 115: 49-63.

Kidwell MG, Lisch DR, 2000. Transposable elements and host genome evolution. *Trends in Ecology and Evolution*, 15(3): 95-99.

Kumar A, Bennetzen JL, 1999. Plant retrotransposons. *Annual Review of Genetics*, 33: 479-532.

Llorens C, Futami R, Renaud G, Moya A, 2009. Bioinformatic flowchart and database to investigate the origins and diversity of clan AA peptidases. *Biology Direct*, 4: 3.

Luo GH, Wu M, Han ZJ, Li XH, Qian L, Fang JC, 2012. Molecular cloning and analysis of *piggyBac*-like transposon in *Chilo suppressalis* (Lepidoptera: Pyralidae). *Acta Entomologica Sinica*, 55(7): 763-771. [罗光华, 吴敏, 韩召军, 李晓欢, 钱路, 方继朝, 2012. 二化螟 *piggyBac* 类转座子的克隆与分析. 昆虫学报, 55(7): 763-771]

Luo GH, Wu M, Wang XF, Zhang W, Han ZJ, 2011. A new active *piggyBac*-like element in *Aphis gossypii*. *Insect Science*, 18: 652-662.

Meng XF, Shi M, Chen XX, 2008. Population genetic structure of *Chilo suppressalis* (Walker) (Lepidoptera: Crambidae): strong subdivision in China inferred from microsatellite markers and mtDNA gene sequences. *Molecular Ecology*, 17: 2880-2897.

Miura A, Yonebayashi S, Watanabe K, Toyama T, Shimada H, Kakutani T, 2011. Mobilization of transposons by a mutation abolishing full DNA methylation in *Arabidopsis*. *Nature*, 411: 212-214.

Muszewska A, Hoffman-Sommer M, Grynberg M, 2011. LTR retrotransposons in fungi. *PloS ONE*, 6(12): e29425.

Oliver KR, Greene WK, 2009. Transposable elements-powerful facilitators of evolution. *BioEssays*, 31: 703-714.

Prathima L, 2012. An analysis of substitution, deletion and insertion mutations in cancer genes. *Nucleic Acids Research*, 40(14): 6401-6413.

Wang JJ, Du YZ, Wang SZ, Brown SJ, Park Y, 2008. Large diversity of the *piggyBac*-like elements in the genome of *Tribolium castaneum*. *Insect Biochemistry and Molecular Biology*, 38: 490-498.

Weil C, Martienssen R, 2008. Epigenetic interactions between transposons and genes: lessons from plants. *Current Opinion in Genetics & Development*, 18: 188-192.

Zhang B, Chen B, Zhang QW, Kang L, 2009. Insect transposable elements and their functions. *Chinese Bulletin of Entomology*, 46(2): 187-195. [张博, 陈兵, 张青文, 康乐, 2009. 昆虫的转座子及其功能. 昆虫知识, 46(2): 187-195]

（责任编辑：赵利辉）